File Copy

CODIB-D-111/1.5/4
11 February 1966

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

Final Report: Task Team V (Biographics)

Attached for coordination within member agencies and discussion

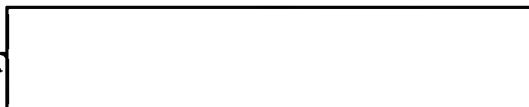at a subsequent meeting is the Task Team V report.

<div style="border:1px solid">
</div>

Secretary

25X1A

Attachment

GROUP I
Excluded from automatic
downgrading and
declassification

25X1

SECRET

SECRET

25X1

U N I T E D S T A T E S I N T E L L I G E N C E B O A R D

COMMITTEE ON DOCUMENTATION

TASK TEAM V -- BIOGRAPHICS

FINAL REPORT

T/V/R-1
1 February 1966

Group 1
Excluded from automatic
downgrading and
declassification. 25X1

SECRET

SECRET [   ]   25X1

T/V/R-1
1 February 1966

UNITED   STATES   INTELLIGENCE   BOARD

COMMITTEE ON DOCUMENTATION

TASK TEAM V - BIOGRAPHICS

MEMORANDUM FOR: Chairman, Committee on Documentation

SUBJECT: Report of Task Team V

1.   Attached is the report of Task Team V for your consideration.

2.   The Team has attempted, in an evolving interpretation of its Terms of Reference, to present realistic recommendations while developing in some depth a substantive description of the problems for the use of interested agencies. While the overall report is classified SECRET [   ] Annex 2 has been given a lower classification to permit wider distribution to U. S. Government officials.

3.   A large file of information, monographs on various aspects of the problem (National Agency Check System, search strategies, data conversion techniques and experienced costs, SCIPS studies of PI files, etc.) is available in or through the CODIB Support Staff.

4.   It is recommended that the Task Team be discharged on CODIB acceptance of this report. A formal mechanism for continued exchange on biographic problems and techniques is, however, contained in the RECOMMENDATIONS.

5.   My thanks to [   ] for his extensive and imaginative support.

[   ]
Chairman, Task Team V

Attachment:
   Task Team V Report

Group I
Excluded from automatic
   downgrading and
   declassification.

SECRET [   ]

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

TASK TEAM V -- BIOGRAPHICS

Table of Contents

Annexes:

1. Glossary

2. Proposed Approach to the Machine Recording of Personal Names
   Attachment 1: Machine Recording Techniques for Personal Names

3. Biographic Index, Facts Summary

4. Data elements in Team Member Agency Records

5. Examples of Name Variants

6. Terms of Reference

7. List of Task Team Members

SECRET   25X1

SECRET 25X1

T/V/R-1
1 February 1966

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

TASK TEAM V - BIOGRAPHICS

FINAL REPORT

PURPOSE

The objective of this Team was to "identify means for improving the storage, retrieval and exchange of information from the major name files and related data files in the Intelligence Community."

SUMMARY OF FINDINGS

1. Improvements in the speed and quality of biographic information processing involving interagency exchange on U. S. citizens and foreign nationals are necessary to further improve security, and to afford policy makers and analysts better response from biographic intelligence files on foreign nationals of interest from a variety of angles--military, subversive, political and scientific. The Team finds that use of computer techniques and inter-agency telecommunications links may provide significant improvements.

2. There are, however, profound, complex problems and significant costs in making major changes in the large biographic holdings of community concern, particularly if the changes involve conversion to computer systems.

3. There are three basically separate, but somewhat over-lapping biographic areas: Counterintelligence* (CI), Positive Intelligence* (PI), and Security*. Name finding* and name searching* take place in all three. (See Annex 1, Glossary, for definition of these and subsequent asterisked terms).

4. The major indexes* considered by the Team ranged from 300,000 unit records (Secret Service) to 50,000,000 (FBI). These now total about 170,000,000 unit records of interagency concern, and are growing at the rate of over eleven million yearly. (See Annex 3).

5. An average of 30,000 requests concerning individuals are made against these indexes daily. Of the 30,000 requests, about

Group I
Excluded from automatic
downgrading and
declassification.

one-half are made between agencies (see footnote) and the other half are processed within the agencies where the requests originate. The 30,000 requests, plus file maintenance procedures, generate 155,000 name searches each day. About one-half of the 15,000 requests made daily between agencies result in a no-record* response.

6. There are several thousand people involved in biographic activity in the Intelligence Community. Approximately 1000 of these, at an annual salary-only cost of $5,000,000 are directly involved, at the index level, in the preparation, maintenance, and searching of the major biographic indexes. These indexes occupy about 100,000 square feet and about $500,000 a year is being spent on supplies and equipment for their support.

7. Agencies in the Washington area are answering security name check requests from each other within two to eighteen days, portal-to-portal, with an overall average response time of nine calendar days. Considerable additional time and cost is involved in delivering the results to the original requester within the requesting agency. The timeliness of response is believed to vary widely owing to volume, personnel costs, and a combination of many other factors unique to each agency. It is difficult to measure the actual loss to the government in terms of personnel not taken on board, personnel taken on board waiting for appropriate clearances, personnel not utilized in a contact or contractual sense because of the slowness of the system. These are intangibles that only the various elements of the respective agencies can weigh within the purview of their own responsibilities and requirements.

8. In the area of name searching, significant quality and time improvements may be obtained through automation and use of tele-communication links. No major name index in the intelligence community has yet been fully automated. Therefore, proof of success has not been conclusively demonstrated. Several agencies are at various stages in developing systems with practical appli-cations anticipated in the near future.

9. The critical problem in any large name index used for name searching is the way in which personal names are recorded, filed, and searched. Any planning for index mechanization must emphasize this aspect. The success of an improved interagency name

---

Note: Since these statistics were gathered, the number of inter-agency name requests submitted by several agencies has increased on the order of 50% during the last several months mainly as a result of several new programs.

SECRET [   ]   25X1

- 3 -

check exchange system based on telecommunications coupled with computer search <u>requires</u> a common approach to recording personal names and certain additional basic identifying data.

10. <u>Name Finding</u> activities could be improved through increased understanding resulting from the exchange between agencies (at both the user and system planning levels) of information about the nature and purpose of each other's specialized files as well as the exchange of data files in certain cases and interchange of information on manual and ADP techniques for improving speed and flexibility of response.

11. The team agreed that the professional interchange derived from the Task Team effort was highly valuable to each member in providing new insights in manual and machine techniques, inter-agency channels, sources of information, and policies of other agencies.

SECRET

25X1

- 4 -

RECOMMENDATIONS

IT IS RECOMMENDED THAT:

1. USIB urge those agencies with large name indexes used for name searching in the National Agency Check system and in Positive Intelligence applications of Community interest to continue to strive within their organizations for index mechanization wherever it is found to be feasible and practical (recognizing that several agencies are already in various steps of development in this area). The findings and report of this Task Team should be used as a point of departure.

2. In conjunction with Recommendation 1, USIB request each agency to study the feasibility of establishing telecommunications links within the National Agency Check complex to facilitate the exchange of requests and replies.

3. USIB request those agencies engaged principally in Positive Intelligence activities to study the feasibility of tying into the Washington area LDX system for the exchange of Positive biographic intelligence.

4. Those agencies which plan to convert large manual biographic indexes to computer-based name searching systems consider the approach to the machine recording of personal names outlined in Annex 2.

5. The CODIB Support Staff be directed to prepare and maintain current publications to inform users of biographic information in the community of the characteristics of each major collection, and the procedures and channels for getting service from each, within the limits of security classification and need-to-know prescribed by each agency.

6. The CODIB Support Staff also serve as the vehicle for informing those agencies developing new computer data files, particularly in the PI biographic area, of the format and coverage requirements of others in the community to reduce unnecessary duplication and coverage gaps.

7. DIA expand its program for the processing of military personality information to meet the needs of the PI community. This should include the processing of open source material and should provide for an EDP file of personality information as well as hard copy backup for such a file. This can be coordinated by DIA with a

SECRET

25X1

group composed of representatives of NSA, CIA, State and cognizant service branches.

8.   The Task Team III (or its successor) be tasked to study those various programs exploiting open source scientific and technical information, which generate personality information of positive intelligence value as a by-product. In conjunction therewith, a coordinated program should be developed using EDP methods to provide machine indexes of the bibliographic data processed by any organization in this field, so that the personality information is accessible to a recipient in machine form, with quick follow-up to the translated source.

9.   Two or three day seminars be held semi-annually (with chairmen rotating from the respective agencies) on the progress of the various agencies in the biographic field, with working sessions for groups with specific problems (such as CI, Security, PI, Communications, the state of relevant technology, software, control techniques, and other functional or technical aspects).

THE NATURE OF THE PROBLEM

1. The Intelligence Community has for many years collected an
ever-increasing amount of information about individuals from a great
diversity of sources through a large number of channels, and has
stored this data in a variety of retrieval systems in diverse formats.
These have traditionally taken the form of index references, either
self-contained or leading to dossier files or individual documents.
The Team decided, as a point of departure, that the relative pay-off
in system improvement would be higher in respect to the larger
biographic files in which there is a high degree of activity and
interagency communication. Thus, many of the smaller files studied
by SCIPS (the Staff for the Community Information Processing Study)
were not included.

2. There are three types of major biographic indexes and files
now in operation. They are the Positive Intelligence, Counterintelli-
gence and Security holdings. There is relatively little exchange of
requests between the PI biographic files and the Security files,
moderate exchange between the CI and PI communities and frequent
exchange between Security and CI. The Counterintelligence (CI)
biographic system centers around the foreign counterintelligence
repository of CIA and the domestic counterintelligence holdings of
the FBI. The security and PI holdings of the agencies referred to
in this report also lead to CI data in some degree. The interagency
exchange of Security data centers around the name search type
operations performed by CIA, State, Army, Navy, Air Force, FBI,
Secret Service, Immigration and Naturalization Service (INS), and
Civil Service Commission (CSC). The major PI biographic records
are contained in the files of the CIA Biographic and Special
Registers, DIA, NSA/Office of Central Reference, Department of State
and Air Force/Foreign Technology Division (FTD).

3. There are important and fundamental differences between, and
some similarities in, the basic operating procedures and kinds of
searches that are made in the PI systems versus the CI/Security
systems. The PI biographic systems are deeply intertwined with,
and in many cases actually part of, larger intelligence collection
and storage systems which are mission, subject or area oriented.
In contrast, the CI/Security systems are clearly oriented to the
heavy use of name searching among alphabetically ordered biographic
indexes which, in most cases, lead to dossier files. The Team
determined that there is name searching and name finding going on in
both the Positive as well as the CI/Security activity. However, the
bulk of the requests in both areas involve name searching (above 95%
in the CI/Security area and about 80% in the PI area).

SECRET

25X1

4. The critical problem in name searching large manual or machine indexes involves the ways in which personal names are reported and stored for retrieval. This is a spelling phenomenon, particularly in PI and CI indexes, which may be classified in two parts:

> a. Name Variants: Different spellings of the <u>phonetically</u> same surname in the original language (SCHUKOW, CHOUKOV, DIUKOV, DZHUGOV, JOUKOFF, YOUKOV, ZHJUKOV, ZHUKOV, etc.). Given name equivalents, diminutives or abbreviations are also considered part of the name variant problem (WILLIAM, WILHELM, WILL, BILL, WM.).

> b. Name Variations: Different conventions in recording and using parts of names (name elements), for example: Fidel CASTRO; CASTRO, Fidel A.; CASTRO y RUZ, Fidel Alejandro; John Taylor BROWN; BROWN, J. Taylor; BROWN, John T.

5. The difficulties in handling the name variant/variation combinations are particularly crucial in those systems in which the preponderance of names are on foreign nationals, or U. S. citizens where control of the source reporting (e.g., employee applications, identification of individual by social security or other number, etc.), is not available. The reasons for the corruption of name spellings received by the majority of agencies considered in this report reflect the real world of intelligence biographics - foreign and domestic. The causes include different transliteration systems between countries (and even within a given country), usage and custom, mistranscription in rewriting names, typographical error, telegraphic garble, and phonetic renditions of names overheard. Examples of these problems are given in Annex 5.

6. Given this situation, the possible combinations and permutations of name variants/variations are unlimited and, more to the point, unpredictable. Thus no formal linguistically based system for reducing name variants to a common denominator has been found wholly adequate for reliable storage and search by those agencies dealing primarily with uncontrolled sources. A pragmatic approach to this problem - called name grouping - is being developed. See Annex 5.

7. The problem is minimized for those agencies which have numerical identifiers (such as social security number or date of birth) in the large majority of their index records. The name variant problem cannot be escaped even so, since these agencies are recipients of name search requests on foreign nationals or U. S. citizens on whom the requesting agency has no control number, and quite possibly a different spelling of the name.

25X1

- 8 -

8. The high proportion of common names adds to the difficulties in large indexes, foreign and domestic. For example, in one multi-million card file on Soviets containing over 300,000 different surname spellings, some 1,500 common surnames account for over 50% of the file. In the case of Vietnam, 54% of the people in the Red River Delta area have the surname NGUYEN; 85% of the Vietnamese population is represented by twelve surnames, with the balance less than 300 clan names.

9. The lack of identifying data on named persons is intimately related to the name variant and common name problems for those agencies without source control. While Annex 4 shows the categories of identifying data recorded if available in the reporting, most foreign and domestic reporting deals with vaguely identified personalities. It is therefore impossible to develop rigid rules on what constitutes the minimum identifying data required. Each agency, in recognizing these problems and the nature of its own index, forms its own rules regarding minimum identifying data for recording, and the depth of search according to the nature of the request.

10. The above indicates what is involved in the quality of name searching. In the past, many agencies have reduced their capability for quality search in manual or machine systems (e.g., by restricting the amount of data recorded). All involved in this Task Team recognize the need to observe the following principles:

    a. Preserve complete name spellings, and record name element components in a consistent format for either manual or potentially mechanized indexes. If an agency is planning the latter, the methodology for the formatting of individual name elements as explained in Annex 2 should be considered.

    b. Retain in the index record all identifying data which assists in distinguishing persons of the same or similar name from one another. Such data elements as sex, date of birth, place of birth, citizenship/nationality, occupation/profession, location, social security number are generally agreed to be desirable, if available, though additional amplifying data further distinguishing the individual should be recorded - regardless of the feasibility of machine search - for human analysis.

    c. Follow the progress of the "name grouping" approach to the name variant problem and, should it prove operationally successful, take advantage of already developed computer techniques to capitalize on the linguistic effort expended by the Government and private agencies for this purpose.

25X1

11.   It was also found that name finding requires substantially
more time and effort per search. This is true because a name finding
request generally must be structured in a more complex fashion and
requires a more involved search procedure.

12.   The Team decided to consider the CI and Security systems
as one area and the PI biographic systems as a separate area for the
purposes of developing the facts, defining the problems, and making
recommendations in this report.

COUNTERINTELLIGENCE AND SECURITY

1.   The Security activity clearly stands out as a network of ten large indexes which are heavily used. Name searches are conducted mainly for granting security clearances for a variety of reasons such as employ-ment, contact, association, contract, etc., and at a variety of security levels. An agency's requirement to grant such a clearance results in the selective checking by that agency of an average of seven other agencies. The major agencies involved in this program include CIA, State, Army, Navy, Air Force, NSA, FBI, Immigration and Naturalization Service, Secret Service, and the Civil Service Commission. The latter three listed are not part of the USIB Community but, in formulating the Team, it was recognized that these agencies are an integral and significant part of the National Agency Check (NAC) Program. Of the approximately 114 million unit records in the Security holdings, these three agencies hold approximately 50 million (I&NS, 37 million; CSC, 12 million; Secret Service, .3 million). Of the 28,000 requests generated daily in the CI/Security System, approximately 8,000 are generated by these three agencies.

2.   Intertwined with the Security request activity are the foreign and domestic Counterintelligence activities centered respectively in CIA and FBI. There are, however, some CI functions in most of the other agencies represented. The normal purpose of the Counterintelli-gence biographic name check activity, as it takes place between the agencies, is to determine the presence of information about an individual of interest to the requesting agency for some counterintelligence reason (e.g., relating to hostile activities of foreign intelligence services and the Communist Party). The CIA maintains a separate and significantly large foreign counterintelligence index [   ] in light of its foreign counterintelligence responsibilities under NSCID 5/3. Security indexes lead primarily to investigative cases and criminal records, predominantly on U. S. citizens. In spite of the fact that requests are made of the CI/Security holdings for different reasons, the nature of the requests and the structure of the data bases involved are substantially the same.

25X1B

3.   The various contributing agencies are listed in Annex 3 along with a set of facts about the respective size, type, growth, activity, etc., of their CI/Security files. It can readily be seen that the size of the various indexes ranges from 300,000 in the case of the Secret Service to over 50 million in the case of the FBI. Most of the unit records are still on 3 x 5 cards. Some of the individual agencies are in the process of converting their indexes to machine language at the present time. This is true of the Office of Security and the Clandestine

SECRET ⬜ 25X1

- 11 -

Services of CIA. The Army and Navy indexes are already on IBM cards, and the NSA Security Records are on magnetic tape. As a result of recent DoD action, the Army, Navy, and Air Force are completing plans to merge their three index holdings on punched cards by mid-1966. Consequently, the Air Force will shortly convert its 3 x 5 index cards to IBM cards for insertion into the common DoD index. This DoD index, although to be in machine language (IBM cards), will, in its initial phase of development, be searched manually. The Immigration and Naturalization Service is presently studying a program to convert its index to machine language and prepare for a machine-based system. This is likewise true of the Secret Service, FBI, and the Civil Service Commission.

4. The CI/Security indexes are growing at approximately 7% per year. This means that they will double in size within ten years at the present rate of growth. Of particular significance is the fact that the 28,000 requests made per day in these indexes (along with the daily maintenance) results in over 120,000 actual name searches being made, mostly manually, in these indexes each day. Of these 28,000 requests, approximately half are made between agencies. From these 14,000 name checks flowing between the agencies, more than half result in a no-record response by the responding agency.

5. The elements of the CI/Security search process considered by the Team include the size and the activity between the agencies, the accuracy and form of the requests and responses, as well as the time that it takes the agencies to respond to each others' requests. The Team noted the fact that there are literally dozens of name check request forms now being utilized by the various agencies. In observing some of these typical and most widely used forms, the Team found that certain basic data such as name, place and date of birth, service serial number, social security number, sex, etc. were included on each form. The Team considered a study of the need for a single name check form to be used by the various agencies. It was considered more important, however, to examine the data elements used and what rules should be applied to their control. These considerations become increasingly critical as the agencies move toward greater use of machine language.

6. To obtain a reasonably dependable determination of the kind of response time in which the various agencies were providing information to each other, a sample survey was made of 3,000 individual typical routine requests. Emergency and priority requests are handled by every agency in a matter of minutes or hours depending upon the results of search. The FIB, I&NS, CSC, CIA, and Army participated in this test.

SECRET ⬜ 25X1

SECRET

25X1

- 12 -

These agencies tabulated the response times of requests from each
other as well as from the Navy and the Air Force. The interagency
response time varies from two to eighteen days with the average of
all the agencies being nine calendar days. There were factors which
the Team recognized as causing possible aberration in these figures:
hand carrying of the requests by liaison personnel, the variations
in the depth of searches, (i.e., on the head* or checking different
possible spellings of the same name) and the researching of the
files by requesting agency personnel on the premises of the answering
agency. In spite of these, the Team feels that the nine-day figure
is a reasonably accurate estimate of the average time (within a day
or two) required for processing of the great bulk of the name checks
being made in this system.

7. It should be noted that the response time referred to
above does <u>not</u> include any <u>internal processing time</u>, in or out, by
the various requesting agencies. The time was measured in all cases
from the day the request left the requesting agency to the day that it
returned to the requesting agency. This time included the mail time
plus that required to make the index search by the responding agency
and the analysis of files in the case of possible identification.
Based on informal observations of the various Team members it appears
that, in the great majority of these cases, there is far <u>more</u> time
spent processing these requests <u>within</u> the requesting agencies
(i.e., from the time the original requester - e.g., analyst, investi-
gator, Ambassador, etc. - sends out his query to the point where it
re-enters the agency and is provided to the ultimate user) than the
nine-day figure of external processing time explained above. To
determine the extent of the internal processing lags and the reasons
therefor was a task far beyond the capability of the Team.

8. Many CI requests are answered from materials that are not
processed into the files, such as directories, working aids, etc.,
or from material too current to be in the file, such as today's
newspaper. Some files are restricted by security classification as
to what can be processed. Research in such a limited source file
often gives incomplete or out-dated information. It is doubtful
that any single file, whether it be computerized or manual, can
ever be considered a complete or sole source for biographic information.

9. It was not possible for the Team to consider specifically the
relative merits of: (a) the improvement of the manual systems within
each agency, (b) the potentials in automation of the index systems
within each agency, and (c) the system efficiency that might be
realized by the institution of a machine language communication system
between the various agencies. These are tasks requiring management
supported feasibility studies, dominated by the professionals within
each agency, in terms of the unique history and problems of each.

SECRET

25X1

POSITIVE INTELLIGENCE

1. The positive intelligence (PI) biographic files can be defined as those files in the intelligence community that have been developed to support the evaluation and production of foreign intelligence. The files are used primarily by government reports officers, researchers and policy makers in establishing or determining facts and reaching decisions in the fields of foreign affairs and defense. The personalities contained in the community's PI files are predominantly foreign nationals. The team concentrated its review upon the major files of the PI community (see Annex 3) on the assumption that the problems involved in the areas of storage, retrieval and exchange would also exist in other PI files and because a large number of the smaller subject-oriented PI files contain the same source material. Development of these smaller files may often be the result of the problems of size, immobility and accessibility that have developed over the years in the large PI files.

2. The management of a PI file can be broken down into four functional areas: collection of source material, selection of information for the files from the source material collected, processing of information into the files, and dissemination of information from the files. The task team concentrated mainly on the area of dissemination and procedures for searching information requests. Since the other three areas have a definite effect on dissemination, they were reviewed.

    a. Collection - Literally hundreds of thousands of source documents are received by a PI file system each year. They will be in English or in a foreign language and each must be read and evaluated. These sources will include the following: newspapers, press services, foreign journals, books, government publications, radio broadcast information and the entire intelligence output of the US intelligence community. A portion of this material will be of a very current nature, having been produced the same day or the previous day.

    b. Selection - The basic criterion of any agency for selecting an item for a PI file is whether or not the item supports the foreign intelligence effort on a particular country or area. Every organization has its own standards for selection based on the mission it is supporting and budgetary limitations. The same source document is frequently processed by different PI organizations. The amount of information that is already available in authoritative sources such as military registers, directories, etc., will often determine what will be

selected for the files. On areas such as the USSR, China, etc.,
a great deal of open source and classified intelligence will be
processed because reliable directory type information is not
obtainable. There is an overlap of information in PI files
because the different file systems support the same requirements,
or because the personality mentioned in the source report meets
the selection criteria for two different requirements: e.g.,
CIA and State have an interest in military personalities who
are prominent in other fields such as politics, science, space,
etc., whereas DIA and NSA are interested in the same person be-
cause he is in the military field. There is no assurance,
however, that a personality mentioned in a source document will
necessarily be processed into a PI file.

c. Processing - Most PI organizations process an abstract,
page or the entire document into its file. The main file may
be in the form of a dossier or a structured alphabetical file
which can be approached directly or through a card or machine
index. The file items may be photocopy, microfilm, multilith,
typed abstract, or the original document. Because of the
timeliness of some information (the same day or previous day)
and the current nature of some requests, it is necessary
either to process this information on a priority basis and get
it into the file quickly or to arrange support files that will
give a researcher quick access to this information. The file
item may be indexed for a particular computer file at the
same time it is processed into a manual PI file system. The
personality name as it appears in a source document is often
either incomplete or misspelled and the name is researched
and corrected wherever possible. Routine processing time from
selection of an item to filing the item will range from an
average of seven to twenty days.

d. Dissemination - The dissemination of information from
a PI file will be usually one of two types: the ad hoc research
of a specific request for information on personalities or the
production of biographic intelligence by the PI element itself.
Examples of the latter are the biographic handbooks produced by
CIA and DIA on high level personalities, Soviet Men of
Science, Biographic Briefs, and the Directory of Soviets.

3. In order to analyze the biographic request activity, the
team members from DIA, CIA, NSA, and State each exchanged a group of
typical research requests. These requests could be grouped into the
following categories: diplomatic and government; military; scientific
and technical; subversive; foreign trade; business and international

SECRET

SECRET   25X1

- 15 -

organizations. The requests involved either name searching, where the identification or complete information on a named individual is requested, or name finding, where the name of the person(s) is either missing or so badly misspelled that research on the other data elements available, such as his position, location, organization or persons associated with him, is required.

4. The group arrived at the following conclusions as a result of its analysis of the requests and its discussion and review of the file systems.

a. PI requests are basically 20% name finding and 80% name searching. It takes more time to research a name finding request, particularly if identifying data in the request is incomplete. A name finding request may generate a list of hundreds of personalities of possible relevance. Many name searching requests require the analyst to use various name finding approaches. If the requester wants a complete identification or biographic sketch on a person holding a government position or an organizational position, (e.g., commander of the Moscow PVO district, General

25X1 [   ] it is necessary to check the records by organization. This will insure that any documents reflecting his change in the organization by position but not name might provide the desired information.

b. A computer system that is developed to process PI information should provide the researcher with both name-searching and name-finding approaches. In a manual system this is usually accomplished by two file systems: a name file in which the personality is searched by his name, and by files that are set up by the other data elements such as organization, location, occupation, etc. In a computer file of limited size, e.g., one or two magnetic tapes, where the maximum search time is fixed, a single file containing the name and all pertinent data elements may be adequate. This will not be true of a file system containing millions of personality records growing at the rate of a million records per year. If name finding approaches are not provided in a large system, the result may well be the development of a new group of subject-oriented files, either manual or computerized, similar to those that presently exist, to meet the needs of specific components of an organization.

c. Many PI requests are answered from materials that are not processed into the files, such as directories, working aids, etc., or from material too current to be in the file, such as today's newspaper. Some files are restricted by security classifi-

SECRET   25X1

cation as to what can be processed. Research in such a limited source file often gives incomplete or out-dated information. It is doubtful that any single file, whether it be computerized or manual, can ever be considered a complete or sole source for biographic information.

    d. "On the head" name search (i.e., researching the name only as it is spelled in the request) cannot always be considered adequate in the PI areas. This implies that information will be found under the name spelling in the request; and since PI name spellings do not usually come from official sources, they are more likely to be incorrect than names found in those indexes where source data is controlled. As mentioned previously, an effort is usually made to correct the spelling before an item is filed, and the same effort is and must be made when performing research.

    e. The PI request is often of a current and timely nature, requiring an answer within an hour or even minutes if it is to be useful to the requester. Routine requests are normally answered within a day. Some extensive research projects may involve thousands of names and require weeks or months to complete. The need for rapid response is one of the reasons a PI element often cannot rely on another agency to answer its requests. This is one of the reasons for the overlap found in the various PI files. The present communications between agencies is not adequate for quick exchange of classified information.

    f. There is an extensive but insufficiently coordinated effort in the intelligence community to produce or bring under control scientific information from open sources on the Soviet Union and Eastern European Communist countries. This activity results in the creation of a great deal of personality information on scientists at all levels of significance.

    g. The community could benefit from a coordinated effort in the production of military biographic information from open sources.

    5. It was not possible for the Team to consider specifically the relative mertis of: (a) the improvement of the manual systems within each agency, (b) the potentials in automation of the index systems within each agency, and (c) the system efficiency that might be realized by the institution of a machine language communication system between the various agencies. These are tasks requiring management supported feasibility studies, dominated by the professionals within each agency, in terms of the unique history and problems of each.

GLOSSARY

COUNTERINTELLIGENCE BIOGRAPHIC AREA: That activity which deals with information on personalities who constitute a known or possible threat to national security. These normally include members and agents of foreign intelligence services, Communist Party officials, and others engaged in organized subversive activities.

POSITIVE INTELLIGENCE BIOGRAPHIC AREA: That activity which deals with information on personalities, usually foreign, who are of general interest to the intelligence community. These include leaders in the scientific, political, governmental, economic, military, and other professional/governmental fields.

SECURITY BIOGRAPHIC AREA: That activity which deals with information held by those organizations which have the normal function of investigating and granting clearances on individuals or organizations. This activity includes information of counterintelligence interest in respect to the internal operations of the holding organization.

NAME FINDING: Searching to identify individuals from data elements other than the name, such as age, position, location, organizational affiliation, occupation, military rank, nationality, including a combination of such factors.

NAME SEARCHING: Search of indexes or files organized by the names of persons to determine if information exists on the individual, or to validate basic information.

MAJOR NAME INDEX: Those personality indexes, in or associated with the intelligence community, which are large in size (several hundred thousand or more unit records) and which are regularly consulted on a routine basis by at least several of the intelligence community member agencies.

ON THE HEAD SEARCH: This consists of a name search on the exact spelling given. For example, a request on BURKE, Robert M. results only in a search in the index against the name BURKE, Robert M, and not any variation of the name. This is the strict interpretation, but some groups which operate biographic holdings in the intelligence community indicate that this definition might include, from the example above, such variations as Robert no middle initial; BURKE, Robert Meredith; BURKE, and BURKE, R. M. All are fairly well agreed that it would not include variants of the spelling of BURKE.

NO RECORD RESPONSE: This refers almost exclusively to name searching. This involves the situation where the check being made results in no

information about the individual <u>at the index level</u>. This is the basis
of the statistics as reflected in column 15 of Annex 3. This does not
reflect the situation where several possible identifications are made
at the index level which, when later analyzed from file information,
are determined to be different individuals, in which case a no record
response still is returned to the requesting agency, nor the numerous
cases in which one or more similarly named persons may possibly be
identical with the subject of the request.

ANNEX 2

PROPOSED APPROACH TO THE

MACHINE RECORDING OF PERSONAL NAMES

INTRODUCTION

1. A USIB endorsed approach to the machine recording of personal names is proposed, subject to qualifications outlined below. The purpose in proposing the adoption of this approach is to insure that those agencies automating their indexes for name searching purposes, where continuing inter-agency exchange is involved, recognize the problems of identifying the elements of personal names in machine recording, and adopt similar, if not identical, logic in storing, maintaining and searching these name elements. This is necessary if the agencies concerned are to exchange, eventually, formatted queries via telecommunications facilities, for input to automated biographic indexes with little or no programmed format conversion and manual reprocessing.

2. In suggesting this approach, it is recognized that significant problems could confront those now using or developing manual or EAM indexes. It also is not intended to preclude the immediate adoption of electrical communications between agencies for speedier search request response.

3. The proposed approach is subject to the following qualifications and assumptions:

    a. It is intended to apply only to those major PI, CI, and Security indexes consulted regularly on an inter-agency basis (e.g., Major NAC indexes, Biographic Register, NSA/CREF), though the approach to personal name recording should be of value as well to those developing internally-used index systems.

    b. The approach assumes computer data recording and manipulation, as opposed to punched card systems (the rules can only apply to variable length records and computer programming techniques to manipulate data elements internally).

    c. The proposal assumes that the rules would be applied only at that point when an agency begins machine language preparation of new input for eventual computer operation, and is not intended to apply to existing punched card records which, however imperfect, may be the only means for converting an existing file to a computer data base.

C-O-N-F-I-D-E-N-T-I-A-L

4. It is felt that those agencies contemplating eventual conversion to computer search systems should evaluate the desirability of recording personal name and related identifying data in variable length input format for computer processing. This will accomplish the beginnings of a data base which will not require later keypunch conversion, provide means for manipulating and editing index information not possible in EAM or manual systems, and will provide also the capability to print or punch index records as a byproduct to keep up manual and EAM systems during the interim stages.

5. Attached hereto is a description of machine recording techniques classified FOR OFFICIAL USE ONLY.

DEFINITION OF ELEMENTS ON THE BIOGRAPHIC INDEX FACTS

1. The index size refers to the actual number of index records (3 x 5 cards, IBM cards, logical records on magnetic tape, etc.).

2. The type of index record would include whether it is a 3 x 5 card, 5 x 8 card, IBM card, on magnetic tape (MT) in document form, etc.

3. The increase per year is the best possible estimate of the yearly change in the number of the index records during the next three years.

4. A multiple reference card is one which leads to more than one dossier, document, etc., by some reference mechanism such as a number.

5. The emphasis in this definition is on the word "predominately" with the understanding that probably all indexes being considered are mixed to some degree. The purpose of this item is to indicate in general terms whether an index mainly concerns U. S. citizens or foreign nationals.

6. See Annex 1.

7. A "request" means a requirement levied on the index, either by the organization internally or by another organization, for the checking of a name of a person. If the request is in the form of a list, for example, names of ten different individuals are considered ten requests.

8. The average number of searches per request indicates how many different ways on the average a request is searched. The searcher may look for a variation in the name, for example, E. J. Jones, Ed Jones, etc., or for the name variant in either the surname or other name elements (for example Nicholas, Nichols, Nickols, Nickles, etc). Some organizations may make one or both types of multiple searches on a certain type or percentage of requests.

9. This is the product of column 7 times column 8.

10. Maintenance searches include such activities as prechecks

SECRET   25X1

BIOGRAPHIC INDEX FACTS SUMMARY

Agency	1 Index size (in millions)	2 Type of unit record (3x5, IBM, etc.)	3 Increase per year (in millions)	4 Single or multiple reference	5 Predominately US Foreign or both	6 PI, CI, or Security type index	7 Total number of day requests	8 Average number of name searches per request	9 Name searches made per day from requests	10 Number of searches per day for maintenance	11 Total number of searches per day (9+10)	12 No requests (7) recorded at index level	13 No requests (7) from other agencies	14 Number of requests from other agencies (total)	15 % of external checks that results in NR	16 Number of external checks that are NR
1 CIA-RID																
2 CIA-OS																
3 Army-RC	9.0	IBM	.5	S	U	S,CI	2650	6	5900	3500	15900	40	33	833	54	477
4 Navy-ONI	4.5	IBM	.3	S	U	S,CI	1500	1.5	2250	3000	5250	65	25	375	65	244
5 AR-OSI Dir	.2.5	3x5	.1	SM	U	CI	1970	1.5	2960	500	1970	45	35	690	81	556
6 AF-OSI DO-4	4.5	3x5	.3	SM	U	S,CI	1200	1.5	1800	1500	1200	57	23	276	3	8
7 NSA-OS																
8 FBI	50.5	3x5	1.0	S	U	S,CI	9000	4	36000	8000	44000	60	61	5400	70	3780
9 I&NS	37.0	3x5	3.0	S	F	S	3000	3	9000	14600	23600	27	8	240	13	31
10 CSC	9.5	4x6	.4	S	U	S	3400	1.5	5100	600	5700	70	all		80	
	2.0	4x6	.1				1300	4.	5200	100	5300	80	all	4700	90	3420
11 SecServ.	.3	3x5	-	S	U	CI	3300	2	600	50	650	50	1	NA	NA	NA
12 CIA-SR																
13 State SY/INR	1.5	3x5	1.	S	B	PI,S,CI	315	1.5	470	350	722	50	35	100	80	80
14 CIA-BR																
15 NSA-OCR																
16 DIA	3.0	5x7 Doc MT	.3	SM	F	PI	900	2.	1800	3000	4800	2	90	810	10	2
TOTALS 1-16	171.7		11.28				33462		84755	51100	155227			14924		9436
1-13	137.2		8.62				31250		79395	37500	118707			13764		9399
12-16	41.5		4.26				2587		6010	17350	40822			1263		117

NA=Not ascertainable.   Lines 1-13=CI/Security Systems   Lines 12-16=PI Systems

25X1B

25X1

25X1B

25X1B

25X1

ANNEX 3

for any reason, the filing of new cards, the refiling of cards for any reason, activity involved in correction of cards, cards being placed or removed for the purposes of opening new cases, purging operations and any other index search or look-up which is not made directly as a result of a normal request as defined under item 7.

11. This is the summation of items 9 and 10. This item reflects the actual total number of searches performed by the reporting organization per day.

12. This is the percentage of the requests (item 7) on which no record or no identifiable information is obtained from a check of the index. It was recognized by the Team that many possible identifications made at the index level later result, after final analysis, in a no record or a no identifiable information; but it was agreed by the Team that since this figure was not readily available, the best criterion for the purposes of this report would be the no record at the index level.

13. This percentage figure represents that proportion of total requests (item 7) which come from other agencies.

14. This represents the number of requests from other agencies as calculated from the percentage figure in column 13 times the request figure in column 7.

15. This percentage figure indicates the portion of requests from other agencies for which no record is found at the index level. The same criterion was used as for item 12.

16. This represents the number of external requests on which no record is found at the index level. It was computed from the percentage figure in column 15 times the number of requests in column 14.

FOR OFFICIAL USE ONLY

- 1 -

Annex 2
Attachment 1

MACHINE RECORDING TECHNIQUES FOR PERSONAL NAMES

1. Described below are some of the problems involved in the
recording, filing, and searching of personal names and suggested
solutions. The problems in the handling of personal names by
electronic data processing are dealt with specifically and considera-
tion is limited to large personal name indexes where (1) point of
retrieval is on name spelling, (2) the quality of name recording,
i. e., spelling and/or completeness of name, cannot adequately be
controlled, e.g., names recorded in newspaper articles, heard on
radio broadcasts, copied from documents, or obtained from second
or third hand sources whose knowledge of the name spelling and/or
completeness may not be reliable, and (3) where additional identifying
information such as date and place of birth, occupation, etc., may
not be consistently reported, and such specific numeric controls
as social security number, military service number, drivers registra-
tion number, etc., do not apply. These conditions are found not only
in the names recorded in an index, but also in the names received as
requests for information.

2. The first problem in recording personal names is to define
the basic order in which the name parts will be recorded. That is,
shall the name be recorded in the English signature style (given
names followed by family name) or in telephone book style (family
name followed by given names)? If the index in question stores
names of all nationalities (very few do not), either style of
recording will require some rearrangement of name parts at the time
of recording. For example, Hungarian and Chinese name signatures
are quite different from the English signature style. That is,
the Hungarian or Chinese name is usually written with the family
name first, followed by the given names.

3. Regardless of the recording style selected, it is important
to define various elements within a name and to identify them in some
manner when they are recorded. The definition and identification of
various name elements is necessary to (1) adequately describe
recording rules to reporters and recorders as they apply to names of
various nationalities, (2) facilitate accurate filing of the name
records in the index, (3) permit accurate machine processing
(sorting) for alphabetic listings, etc., (4) and to facilitate storage
and retrieval (search) of name records by computer.

FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

4. Many different codes, symbols, characters, or fielding techniques may be used to identify various name elements. However, if a printed version of the name is to be read by persons not normally associated with the EDP environment, it is preferable to use common punctuation which can easily be interpreted by the customer, i.e., use a period after a single alphabetic character to identify an initial as opposed to a single character name or particle.

5. Definitions of various name elements which should be identified when recording the name follow:

   a. NAME: That word or combination of words used to identify a person.

      (1) The minimum field length for recording the name should be forty characters. Although many names can be recorded in less than 40 characters, the truncation imposed upon lengthy names by, say, a 20 character limit, often eliminates the very elements which provide discreteness. Such system-imposed restraint increases the number of name records which will be retrieved in a search. Additionally, it often imposes pre-input editing to be sure that critical elements of the name can be recorded in the field size allotted. For example, the name Evangelica Concepcion Rodriquez y Gonzalez contains 42 characters including spaces and without any special characters to identify various name elements. The usual pre-input edit of this name would probably reduce it to RODRIQUEZ, EVANGELIC, thus making it impossible to distinguish this Evangelica Rodriquez from any other Evangelica Rodriquez. If the name were not pre-input edited, but merely truncated by the input typist or arbitrarily by the machine, the entry RODRIQUEZ Y GONZALEZ, EVANGELICA CONCEPCION would be truncated to RODRIQUEZ Y GONZALEZ which is even less discrete. Forty characters permits recording of the family name and most of her given names, i.e., RODRIQUEZ Y GONZALEZ, EVANGELICA CONCEPC.

FOR OFFICIAL USE ONLY

b. <u>SURNAME</u>: The word or words which comprise the element of a name commonly referred to as the "last name" or "family name," including initials, abbreviations, and particles (defined below) if reported as part of the surname. <u>The surname is that element of the name which governs the primary position of a name in an alphabetic file.</u> Surnames containing more than one word are referred to as "compound" or "Multi-Word" surnames.

(1) Because surnames often contain more than one word, and in view of its basic importance to the filing and subsequent finding of the name record, it is necessary to identify which part of the complete name is the surname. In the examples which follow, surname is printed first followed by a comma to show the end of the surname. If some such method of surname identification is not used, surnames which contain more than one word cannot be distinguished from those with only one word followed by first name.

Examples: BROWNE, T. R.

                CESPEDA Y LOPEZ, JUAN

                KAMAL AL DIN, MOHAMED

c. <u>GIVEN NAME</u>: The word or words in a name commonly referred to as the "first," "baptismal," "Christian," "middle," or "patronymic," etc. Initials and abbreviations are included. <u>Given Names dictate the alphabetic position of a name record within like surnames.</u> Therefore, particles, titles, and telecodes (defined below) are not included in the definition of "Given Name."

(1) Whether the name parts being recorded are called "Surname and Given Name" or "Clan Names" or whatever, is irrelevant. It is important, however, to identify which word or words in a name are to be used as the primary storage or search element (Surname) and which are to be used secondarily, (Given Name).

(2) Note, in the following list of names recorded without commas, that "compound" surnames cannot be distinguished <u>by a computer</u> from non-compound surnames and, therefore, the second word of the compound surname is likely to be used as a given name.

GARCIA LOPEZ JOSE	should be	GARCIA LOPEZ, JOSE
MAC DONALD HENRY	" "	MAC DONALD, HENRY
RODRIGUEZ L. JUAN	" "	RODRIGUEZ L., JUAN
ST. CLAIR ROMAN LUIS	" "	ST. CLAIR ROMAN, LUIS
STA. ANA RAUL	" "	STA. ANA, RAUL

d. PARTICLES: Particles include the articles (la, der, etc.) prepositions (de, von, etc.) and conjunctions (und, etc.), foreign equivalents of the English the, of, and, etc., which have not become an integrated part of the name.

(1) Particles are usually ignored in the filing of names because they may be different each time a name is reported and recorded or may at times be completely absent. Therefore, if the particles were used in determining the alphabetic file position of the name, the same name would be filed in different places.

Examples: GARCIA LOPEZ, JUAN

GARCIA (Y) LOPEZ, JUAN

GARCIA (E) LOPEZ, JUAN

(DE) GENNARO, GUISEPPE

(DI) GENNARO, GUISEPPE

GENNARO, GUISEPPE

KAMAL (AL) DIN, MOHD

KAMAL (UD) DIN, MOHD

KAMAL (EL) DIN, MOHD

KAMAL (ED) DIN, MOHD

(2) For the above reasons, it is important to identify those words in a name which are particles. When they have been properly identified, the computer

processing of these names will be able to facilitate appropriate alphabetic sequence.

(3) For name searching purposes, it is particularly important that particles appearing in the given name field be identified (for example, by enclosing in parenthesis) so that they are not confused with given names.

Examples: NASSIR, GAMAL ABD (AL)

SHARIF, ABD (AL) MOHD

e. TITLES: A descriptive name or appellation which denotes rank, office, privilege, or is used as a mark of respect. The terms Jr., III., 2nd, Mrs., Miss, Colonel, Prince, etc., are included as titles.

Example: BROWN, JOHN /JR/

(1) In most files dealing with military personalities, rank is normally fielded separately. If titles are included in the name field, it is important that they be identified as such, so that they do not become confused with given names.

Example: SCHEINHEIMER, BARON should be

SCHEINHEIMER, /BARON/

f. TELECODE: Numeric equivalent of ideographs used in Chinese, Korean, and Japanese writings. Some Japanese ideographs which have no numeric equivalent are represented phonetically, i.e., "KATAKANA." When the ideograph is illegible and/or the numeric equivalent is not known, the term, NTA (No Telecode Available) is often used.

Examples: TOJIMA, FUSANOSUKE /2073/*02701/2075/0037/6534/

LEE, WON-LOU /NTA/0029/0283/

CHAN, LI-SHU /7115/0173/0209/

(1) Each numeric or alphabetic set in the telecode should be separated from the other by some special character. If the telecode is recorded in the name field, special characters should be used to identify it for potential special processing by the computer.

FOR OFFICIAL USE ONLY

g. <u>PREPARATION OF THE NAME FOR SORTING AND STORAGE:</u>

(1) If characters other than alphabetic are used in the name, certain special characters should be removed for sorting purposes, creating a so called "Pure Name" for sorting purposes. The internal creation of a sort name is necessary to assure accurate sequencing of names for alphabetic printing or storage. When the name is printed, the original input name field is used.

(2) If characters such as hyphen or an apostrophe were allowed to remain in the name during a sort, the name HERNANDEZ-PELAGIO would be listed after the name HERNANDEZ ZERTUCHE. A search for O'BRIEN would find it listed before names beginning with OA and not in the OB part of the list as would be expected.

(3) Characters and special elements to be removed for sort purposes are:

(a) Particles - remove and left justify the remainder of the name.

(b) Hyphen - remove and insert space.

(c) Period - remove and left justify the remainder of the name.

(d) Comma - remove and insert an extra space code.

(4) Titles and telecodes included in the name field are sorted to numeric and/or alpha order. The virgules or other special characters enclosing these characters are also used in sorting and will provide the uniqueness required to place names embodying titles or telecodes after like names in the file, without a title or telecode.

(5) Upon the removal and substitution of the foregoing, the name may be sorted accurately to alphabetic order. Note, in the following examples, the effect of the foregoing rules, especially with respect to compound names.

NAME AS PRINTED	NAME FOR SORTING
'AZIM, MOHAMED (AL)	AZIM MOHAMED

FOR OFFICIAL USE ONLY

- 7 -

NAME AS PRINTED	NAME FOR SORTING
AZIM, MOHAMED AL	AZIM MOHAMED AL
GARCIA, MARIA	GARCIA MARIA
GARCIA-LOPEZ, MARIA	GARCIA LOPEZ MARIA
GARCIA (Y) LOPEZ, MARIA	GARCIA LOPEZ MARIA
O'BRIEN, JOHN	OBRIEN JOHN
O'BRIEN, JOHN /DR./	OBRIEN JOHN /DR./
(DE) SANTOS, JOSE	SANTOS JOSE
SMITH, J. X.	SMITH J X
SMITH, J. XAVIER	SMITH J XAVIER
SMITH, ZELAYA	SMITH ZELAYA
SMITH-CORONA, JAMES	SMITH CORONA JAMES
STE. ANTON, GREGOR	STE ANTON GREGOR
STE-ANTON, GREGOR	STE ANTON GREGOR

10. The following, in summary, is the approach the Team recommends in the identification of name elements, with examples of the types of punctuation controls which may be used:

    a. Record complete name elements in a consistent order, i.e., surname followed by given names then by telecodes and/or titles.
    Example: CHIANG, KAI-CHEK /1203/0009/7156/

    b. Identify surname elements as opposed to given name elements, i.e., by placing a comma between the two elements.

    Example: DOE, JOHN

    c. Identify particles, i.e., by placing parenthesis around them.
    Example: GARCIA (y) LOPEZ, JOSE

FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY

    d. Identify titles and/or telecodes, i.e., by placing virgules around them.

        Examples: CHAN, WON LI /0148/0029/0173/

                  ROBBINS, CHARLES A. /JR./

    e. Identify initials from one character names, i.e., by terminating them with a period.

        Examples: SMITH, J. L. ARMAND

                  Y, LI CHU (one character surname)

                  SANCHEZ R., JUAN

    f. Allow sufficient space for recording the entire name. Forty (40) positions minimum are recommended.

FOR OFFICIAL USE ONLY

ANNEX 4

DATA ELEMENTS USED IN BIOGRAPHIC SYSTEMS

DATA ELEMENT	CIA-RID	CIA-OS	CIA-BR (IBM)	CIA-BR (5x8)	CIA-SR	State-SY	Army-RC	Navy-ONI	NSA-OCR (DOC)	NSA-OCR (MT)	NSA-OS	AF-OSI Dir	AF-OSI DO-4	I&NS	F B I	Secret Ser	CSC (U)	CSC (S)	D I A
1 NAME	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
a Title	x	x	x	x		x	x		x	x	x				x	x	x		
b Rank or Grade	x	x	x	x			x		x	x	x					x	x		x
c Alias	x	x	x	x			x	x		x	x	x	x	x	x	x	x	x	
d Also known as	x	x	x	x		x	x		x	x	x	x	x	x	x	x		x	
e Maiden	x	x		x					x	x	x	x	x	x	x				
f Pseudonym	x	x	x	x		x			x	x	x			x	x				
g Ming # (telecode)	x	x		x				x		x	x	x	x					x	
2 Date of Birth	x	x	x	x		x	x	x		x	x	x	x	x		x		x	
3 Place of Birth	x	x	x	x			x	x		x	x	x	x	x		x		x	
4 Nationality		x	x	x					x	x	x		x	x				x	
5 Citizenship	x	x	x	x	x	x	x		x	x	x		x	x					
6 Sex	x	x		x	x	x		x		x	x	x	x		x		x		
7 Race				x						x			x	x		x			
8 File number	x	x	x	x		x	x	x		x	x	x	x	x	x	x	x		
9 Social Sec. No.		x				x	x			x	x	x	x		x				
10 Service No.						x	x			x	x			x	x				
11 Residence	x	x		x			x			x	x	x	x	x					
12 Employment	x	x	x	x			x	x	x		x	x							
13 Occupation		x	x	x	x		x	x		x			x	x					
14 Education				x															
15 Personal Description				x						x			x						
16 Spouse Data		x		x						x			x	x					
17 Military Service			x	x			x			x									
18 Language			x	x															
19 Reference						x				x									x
20 Date of Card											x	x		x		x			
21 Document Reference	x			x											x				
22 Document Date	x			x			x												
23 Record Number	x																		
24 Location of File																	x	x	
25 Year Record Created		x	x	x															
26 Added Admin. Elements													x						
27 Localities	x		x	x	x									x					
28 Phone Caller															x				
29 Letter Writer															x				
30 Added Info	x		x	x	x										x				x
31 Type of Case																	x	x	
32 Addresses				x													x	x	
33 Disposition																	x		
34 Remarks	x																		x

(these elements are entered by each agency as available or required)

THE NAME GROUPING APPROACH

1. The Name Grouping approach is designed to insure that a search of a name brings together all references to an individual although his name may have been recorded in various spellings and transliterations. This is accomplished by having linguists (native speakers) examine the name spellings recorded in a particular index <u>in order to put names which belong together phonetically</u> in a group which is then identified by a number. Thus when the index is searched, references recorded on any variant of a surname or given name are brought together through the pre-analysis and grouping by the language expert.

2. The purpose of the technique is to build into a given index system a one-time, professional linguistic analysis of each unique name spelling related to other phonetically identical name spellings on a purely pragmatic basis. That is, name grouping is concerned with the name spellings actually received by an organization, not by rules or theories on how names might have been, or ought to be, spelled. The primary advantage is to avoid a variety of search criteria by various index clerks.

3. Inherent in this technique is the logic for random access storage of biographic records in a computer system. The surnames and given names are used as computer dictionaries (tables) leading to all group index records on a given name variant in one storage area of a random access file.

EXAMPLES OF NAME VARIANTS

VARIANT SPELLINGS OCCURING FROM TRANSLITERATION

ARABIC   ND L YHM

مسيحي ال ربيع

MUHI-AL-DIN
MAHJOEDIN
MAHAYIDEEN
MAHYUDDIN
MHIDINE
MOHAYUDDIN
MOHHDIN
MOHIDEEN
MOYIDEEN
MOHIEDDIN
MUHY-AL-DIN
MUHYI-UD-DIN
  plus 25 more

FAR EAST   劉   = Telecode 0491
LIU = Mandarin
LAU = Cantonese
YU = Korean
RYU = Japanese

OTHERS   WØGE = WOEGE, WERGE
JÄNSEN = JAANSEN
NÕNEN = NOONEN
IANOZZI = JANOZZI, YANOZZI
SNJDER = SNYDER, SNIDER
MENSKJ = MENSKY, MENSKIY
PETROW = PETROV, PETROF
FELDMAN = FELDMAN, FELTMAN, FELDTMAN

FOR OFFICIAL USE ONLY

- 3 -

EXAMPLES OF SURNAME GROUPS

GROUP	NAME	GROUP	NAME
009546	IZJERMAN EISERMANN	001712	MATZGER METZGER MEZHER
002914	CHLADEK HLADEC HLADIC HLADIK		MAETZCHKER METZKER MEZGER
	HLADK HLADEK	002194	SCHUKOW CHOUKHOV DIUKOV
008687	ABOURGELI RUJAYLAH		DZHUGOV JOUKOFF SCHUCHOW
004739	FOGELER VOGELER VOGLER WOEGELER		SHUKHOV YOUKOV YOUKOVA ZHJUKOV ZHUKOV ZHUKOVA

EXAMPLES OF GIVEN NAME GROUPS

GROUP	NAME	GROUP	NAME
Z00007	ABRAHAM BRAHIM EBRAHIM IBRAGIM JBRAHIM	Z00086	EDWARD EDVARD EDOARD EDUARD EDUART EDVART
Z00650	STEPHAN STEVAN STEVEN ISTVAN ETIENNE ESTABAN STEFAN STEFA STEVE STEVO STJEPAN		SEE ALSO: ED. GROUP #Z00002 EDW. GROUP #Z00018 EDWIN EDVIN EDWINS EDVINE SEE ALSO: ED. GROUP #Z00002 EDW. GROUP #Z00018

TERMS OF REFERENCE

A. OBJECTIVE

To identify means for improving the storage; retrieval and exchange of information from the major name files and related data files in the Intelligence Community.

B. FACT FINDING

1. Identify those significant index and related systems leading to biographic information collections in the government which are routinely consulted by intelligence agencies for their security, counterintelligence or foreign (positive) intelligence content.

2. Establish the following facts concerning each of the above.

   a. Size: Number of index records (i.e., extracts of information, such as 3 x 5 cards, punched cards, magnetic tape records, disk records, strip records, etc. normally leading to documents and files), type and size of index records, single or multiple reference.

   b. Emphasis on types of personalities covered: e.g., percentage of foreign vs U. S. citizens, scientists, military political, Communist Party, Maritime, foreign intelligence services, agents, etc. This will include the "name finding" as well as the "name searching" activity.

   c. Number of names searched daily: Percentage of positive and negative responses, depth of search on name variants.

   d. Major requesters; proportion of requests from each.

   e. Methods of communicating requests and responses: Forms, memoranda, teletape, transceiver, data phone; security classification of requests and responses.

   f. Identifying data in conjunction with name normally included in index reference.

   g. General description of input, maintenance and search processing.

   h. Current requirements for submission of requests.

  i. Classification of the index.

C. REVIEW

    1. Examine costs, methodology and prospects for biographic systems now undergoing mechanization.

    2. Identify basic problems to be faced and areas where policy decisions are required by each agency in planning for mechanization.

    3. Identify those areas where format, methodology and equipment compatibility are required or are highly desirable in name searching or finding to obtain optimum speed, quality and economy in automating query and response.

D. RECOMMENDATIONS

    Formulate recommendations for CODIB and USIB approval outlining policy objectives for the Community, with generalized projections of cost, manpower and time required to meet these objectives. Include specific guidelines for agencies to follow in systems planning and development.

S-E-C-R-E-T

TERMS OF REFERENCE

A. OBJECTIVE

To identify means for improving the storage; retrieval and exchange of information from the major name files and related data files in the Intelligence Community.

B. FACT FINDING

1. Identify those significant index and related systems leading to biographic information collections in the government which are routinely consulted by intelligence agencies for their security, counterintelligence or foreign (positive) intelligence content.

2. Establish the following facts concerning each of the above.

a. Size: Number of index records (i.e., extracts of information, such as 3 x 5 cards, punched cards, magnetic tape records, disk records, strip records, etc. normally leading to documents and files), type and size of index records, single. or multiple reference.

b. Emphasis on types of personalities covered: e.g., percentage of foreign vs U. S. citizens, scientists, military political, Communist Party, Maritime, foreign intelligence services, agents, etc. This will include the "name finding" as well as the "name searching" activity.

c. Number of names searched daily: Percentage of positive and negative responses, depth of search on name variants.

d. Major requesters; proportion of requests from each.

e. Methods of communicating requests and responses: Forms, memoranda, teletape, transceiver, data phone; security classification of requests and responses.

f. Identifying data in conjunction with name normally included in index reference.

g. General description of input, maintenance and search processing.

h. Current requirements for submission of requests.

S-E-C-R-E-T

- 2 -

ANNEX 6

    i. Classification of the index.

C. REVIEW

    1. Examine costs, methodology and prospects for biographic systems now undergoing mechanization.

    2. Identify basic problems to be faced and areas where policy decisions are required by each agency in planning for mechanization.

    3. Identify those areas where format, methodology and equipment compatibility are required or are highly desirable in name searching or finding to obtain optimum speed, quality and economy in automating query and response.

D. RECOMMENDATIONS

    Formulate recommendations for CODIB and USIB approval outlining policy objectives for the Community, with generalized projections of cost, manpower and time required to meet these objectives. Include specific guidelines for agencies to follow in systems planning and development.

S-E-C-R-E-T

S-E-C-R-E-T

MEMBERS OF CODIB TASK TEAM V – BIOGRAPHICS

25X1A

CIA
    Mr. ☐ irman
    Mr. ☐ (Alternate)
    Mr. ☐ (Alternate)

DIA
    Mr. ☐

STATE
    Mr. Mitchell Stanley
    Mr. Halvor Eckern (Alternate)

ARMY
    Mr. Paul Anderson

NAVY
    Mr. Marvin E. Van Dera
    Mr. William Urick (Alternate)

25X1A

NSA
☐ (Alternate)

AIR FORCE
    Lt. Col. Edmund M. Manning
    Maj. Russell S. Keen (Alternate)

I&NS
    Mr. John L. Keefe

FBI
    Mr. Earl W. McCoy

SECRET SERVICE
    Mr. Frank G. Stoner

CSC
    Mr. Pearley G. Buck

CODIB Support Staff
25X1A ☐ Secretary

S-E-C-R-E-T

S-E-C-R-E-T

MEMBERS OF CODIB TASK TEAM V - BIOGRAPHICS

25X1A

CIA

[redacted] airman
(Alternate)
(Alternate)

DIA

STATE
    Mr. Mitchell Stanley
    Mr. Halvor Eckern (Alternate)

ARMY
    Mr. Paul Anderson

NAVY
    Mr. Marvin E. Van Dera
    Mr. William Urick (Alternate)

25X1A

NSA
[redacted] (Alternate)

AIR FORCE
    Lt. Col. Edmund M. Manning
    Maj. Russell S. Keen (Alternate)

I&NS
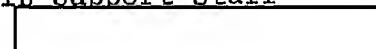   Mr. John L. Keefe

FBI
    Mr. Earl W. McCoy

SECRET SERVICE
    Mr. Frank G. Stoner

CSC
    Mr. Pearley G. Buck

CODIB Support Staff

25X1A

[redacted] Secretary

S-E-C-R-E-T